

[19] 中华人民共和国国家知识产权局



## [12] 发明专利申请公开说明书

[21] 申请号 200410085186.1

[51] Int. Cl.

G10L 13/08 (2006.01)

G10L 13/00 (2006.01)

G06F 17/20 (2006.01)

G06F 17/28 (2006.01)

[43] 公开日 2006 年 4 月 5 日

[11] 公开号 CN 1755796A

[22] 申请日 2004.9.30

[21] 申请号 200410085186.1

[71] 申请人 国际商业机器公司

地址 美国纽约

[72] 发明人 张维 马希俊 金凌 柴海新

[74] 专利代理机构 中国国际贸易促进委员会专利商标事务所

代理人 吴丽丽

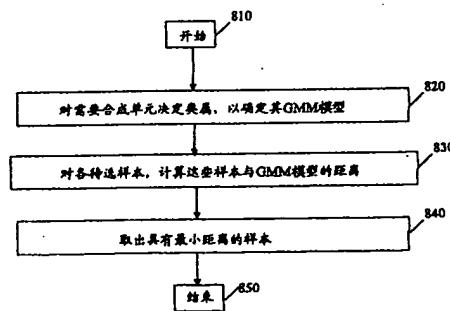
权利要求书 3 页 说明书 11 页 附图 6 页

### [54] 发明名称

文本到语音转换中基于统计技术的距离定义方法和系统

### [57] 摘要

本发明提供一种在 TTS 系统中进行距离定义的方法。该方法将高斯混合模型 GMM 应用于 TTS 的距离定义中，以解决现有 TTS 技术中的诸如数据稀疏和数据分散这样的疑难问题。该方法在合成过程中包括如下步骤：为需要进行 TTS 语音合成的单元决定其类属，由此确定该类属的 GMM 模型；计算该类属中的各个待选样本与确定的 GMM 模型间的距离；取出具有最小距离的那个样本，以用于语音合成。



1. 一种在 TTS 系统中进行距离定义的方法，包括如下步骤：  
对要进行 TTS 的文本进行分析，以获得带有描述性韵律注释的文本；  
针对所获得的文本中的样本进行分类；以及  
为每个分类产生 GMM 模型，确定样本与其相应 GMM 模型的距离。
2. 根据权利要求 1 所述的方法，其中分类步骤包括采用决策树进行分类。
3. 根据权利要求 2 所述的方法，还包括在决策树中的两个分支相似时，将这两个分支合并的步骤。
4. 一种在 TTS 系统中进行距离定义的系统，包括：  
文本分析单元，用于对要进行 TTS 的文本进行分析，以获得带有描述性韵律注释的文本；  
韵律预测单元，用于针对从文本分析单元获得的文本中的样本进行分类，以为所述分类产生 GMM 模型；以及  
GMM 模型库，与所述韵律预测单元相连，用于存储所产生的 GMM 模型。
5. 根据权利要求 4 所述的系统，其中所述韵律预测单元采用决策树来对从文本分析单元获得的文本中的样本进行分类。
6. 根据权利要求 5 所述的系统，还包括合并单元，用于对韵律预测单元所使用的决策树中的相似分支进行合并。
7. 一种在 TTS 系统中进行语音合成的方法，包括如下步骤：  
为需要进行 TTS 语音合成的单元决定其类属，由此确定该类属的 GMM 模型；  
计算该类属中的各个待选样本与确定的 GMM 模型间的距离；  
取出具有最小距离的那个样本，以用于语音合成。
8. 根据权利要求 7 所述的方法，其中所述计算距离的步骤包括

计算目标代价和转换代价。

9. 根据权利要求 7 或 8 所述的方法，其中取出具有最小距离的样本的步骤包括取出目标代价最小的样本的步骤。

10. 根据权利要求 7-9 中任意一项所述的方法，其中取出具有最小距离的样本的步骤包括取出目标代价加转换代价最小的样本的步骤。

11. 根据权利要求 7 所述的方法，其中为需要进行 TTS 语音合成的单元决定其类属，由此确定该类属的 GMM 模型的步骤包括：

    获取每个需要合成的单元的描述性韵律注释信息；

    查找每个需要合成单元的类属，所述类属对应于一个 GMM 模型；

    在 GMM 混合模型序列的空间中，基于距离定义以及总体优化准则搜索最佳值，从而最佳值序列被作为 GMM 模型的明确预测。

12. 根据权利要求 7 所述的方法，其中计算该类属中的各个待选样本与确定的 GMM 模型间的距离以及取出具有最小距离的那个样本以用于语音合成的步骤包括：

    获取每个需要合成单元的描述性韵律注释信息；

    查找每个需要合成单元的属，所述类属对应于一个 GMM 模型；

    通过基于 GMM 的距离定义评估需要合成单元的所有候选；以及

    基于评估步骤中给出的距离以及总体优化准则找出总体优化候选序列，以进行语音合成。

13. 一种在 TTS 系统中进行语音合成的系统，包括：

    类属决定单元，用于为需要进行 TTS 语音合成的单元决定其类属，由此确定该类属的 GMM 模型；

    距离计算单元，用于计算该类属中的各个待选样本与确定的 GMM 模型间的距离；

    优化单元，用于取出具有最小距离的那个样本，以用于语音合成。

14. 根据权利要求 13 所述的系统，其中所述距离计算单元还包括

目标代价计算单元和转换代价计算单元。

15. 根据权利要求 13 或 14 所述的系统，其中优化单元从目标代价计算单元取出目标代价最小的样本。

16. 根据权利要求 13-15 中任意一项所述的系统，其中优化单元从目标代价计算单元和转换代价计算单元二者中分别取出目标代价加转换代价最小的样本。

17. 根据权利要求 13 所述的系统，其中所述类属决定单元包括：

用于获取每个需要合成的单元的描述性韵律注释信息的装置；

用于查找每个需要合成单元的类属的装置，所述类属对应于一个 GMM 模型；

用于在 GMM 混合模型序列的空间中，基于距离定义以及总体优化准则搜索最佳值，从而最佳值序列被作为 GMM 模型的明确预测的装置。

18. 根据权利要求 13 所述的系统，其中距离计算单元包括：

用于获取每个需要合成单元的描述性韵律注释信息的装置；

用于查找每个需要合成单元的属，所述类属对应于一个 GMM 模型的装置；以及

用于通过基于 GMM 的距离定义评估需要合成单元的所有候选的装置；以及

优化单元还包括：

用于基于评估步骤中给出的距离以及总体优化准则找出总体优化候选序列，以进行语音合成的装置。

## 文本到语音转换中基于统计技术 的距离定义方法和系统

### 技术领域

本发明涉及文本到语音转换 (text-to-speech: TTS) 领域。更具体地说，本发明涉及用于文本到语音转换中基于统计技术的距离定义的系统和方法。

### 背景技术

所谓文本到语音转换 (TTS)，是指在计算机的支持下，通过复杂的自然语言处理算法的设计，把文字智能地转化为自然语音流的技术，该技术可以使用户方便地与计算机进行交流，从而大大提高应用系统的灵活性。

如图 1 所示，一个典型的 TTS 系统 100 通常包括三个组件：文本分析单元 101，韵律预测单元 102 以及语音合成单元 103。其中，文本分析单元 101 负责将输入的纯文本分析为具有描述性韵律注释（例如，发音、重音、短语边界和停顿）的带标记文本。韵律预测单元 102 负责根据文本分析的结果预测韵律的发音表示，例如，音调值，每个合成段的持续时间和能量。而语音合成单元 103 负责生成可理解的语言，作为纯文本中隐含的语义和韵律信息的表示的物理结果。

例如，将文本“这是一个专利申请”进行文本到语音转换，将会有下列结果。

首先，将该文本输入文本分析单元 101，于是每个字的发音以及短语边界等信息被辨别出来：

这 | 是 | 一 | 个 | 专 | 利 | 申 | 请。

zhe4 shi4 yi2 ge4 zhuan1 li4 shen1 qing3

在进行了上述文本分析后，韵律预测单元 102 会对文本中的字进行发音预测。

在对文本进行了上述预测后，所述语音合成单元 103 会根据预测的韵律信息生成该段文本的相应语音。

在目前的 TTS 中，使用基于统计的方法是一种重要的趋势。在这些方法中，从大型的有标记的语料库（labeled corpus）中训练韵律预测模型，并且语音合成总是基于每个合成段的多个候选的选择。图 2 示出了通用的基于语料库的 TTS 的框架结构。

在基于统计的方法中，特别是在韵律预测和基于库（inventory）的选择中，很多问题归根到底都是样本和给定类之间的距离定义的问题。这种距离定义不是很容易实现，因为即使类数据有复杂的上下文，但数据分散问题在几乎每个类中还都会严重，并且类之间的重叠也很严重，以致难以评估样本是否属于给定的类。

在目前的 TTS 中有一些经典的基于统计的距离定义方式，如加权 Euclid 距离和 Mahalanobis 距离。但是，它们分别都存在一些问题。对于 Euclid 距离而言，由于其使用各样本的平均值作为样本点，因而通常不能选择最适当的那个值来作为样本点，且不同维度（dimension）间的关系被忽略或通过预先给定的知识被很差地建模。而 Mahalanobis 距离的不足之处在于其不能模拟复杂分布。

图 3 示出了一个直方图，其表示在一个 TTS 语料库的一个类中某个样本的持续时间分布为对数分布时的情形。在图 3 所示的情形中，数据比较分散，因此采用 Euclid 距离定义的取平均值的方式根本无法模拟其分布，又由于该分布不是正态分布，也难以采用 Mahalanobis 距离定义对其进行精细模拟。

### 发明内容

为了克服如上所述的问题，本发明提出将高斯混合模型（Gaussian Mixture Model: GMM）应用于 TTS 的距离定义。具体

而言，本发明提出了一种用于文本到语音转换的统计技术的全新距离定义方式。在根据本发明的距离定义中，通过 GMM 来引入概率分布。由于概率分布的使用，与前述加权 Euclid 距离定义和 Mahalanobis 距离定义相比，能够很好地解决 TTS 统计技术中的诸如数据稀疏和数据分散这样的疑难问题。GMM 是用于通过具有用于每个高斯模型的简单参数的高斯类属来描述某些复杂分布的通用算法。例如，图 3 的分布可由 GMM 组合两个高斯模型来模拟，图 4 是模拟的示意图。但是，本领域技术人员能够理解，用 2 个高斯分布来模拟图 3 的分布只是示例性的，也可以根据情形需要用其他数量的高斯分布进行模拟。

为了实现上述目的，根据本发明的第一方面，提供了一种在 TTS 系统中进行距离定义的方法，包括如下步骤：对要进行 TTS 的文本进行分析，以获得带有描述性韵律注释的文本；针对所获得的文本中的样本进行分类；以及为每个分类产生 GMM 模型，确定样本与其相应 GMM 模型的距离。

根据本发明的第二方面，提供了一种在 TTS 系统中进行距离定义的系统，包括：文本分析单元，用于对要进行 TTS 的文本进行分析，以获得带有描述性韵律注释的文本；韵律预测单元，用于针对从文本分析单元获得的文本中的样本进行分类，以为所述分类产生 GMM 模型；以及 GMM 模型库，与所述韵律预测单元相连，用于存储所产生的 GMM 模型。

本发明的第一和第二方面涉及利用语料库对 GMM 模型进行训练。

根据本发明的第三方面，提供了一种在 TTS 系统中进行语音合成的方法，包括如下步骤：为需要进行 TTS 语音合成的单元决定其类属，由此确定该类属的 GMM 模型；计算该类属中的各个待选样本与确定的 GMM 模型间的距离；取出具有最小距离的那个样本，以用于语音合成。

根据本发明的第四方面，提供了一种在 TTS 系统中进行语音合成的系统，包括：类属决定单元，用于为需要进行 TTS 语音合成的单

元决定其类属，由此确定该类属的 GMM 模型；距离计算单元，用于计算该类属中的各个待选样本与确定的 GMM 模型间的距离；以及优化单元，用于取出具有最小距离的那个样本，以用于语音合成。

本发明的第三和第四方面涉及利用 GMM 模型进行语音合成。

#### 附图说明

图 1 是典型 TTS 系统的框图；

图 2 是通用的基于语料库的 TTS 的框图；

图 3 表示在一个 TTS 语料库的一个类中某个样本的持续时间分布为对数分布时的情形；

图 4 表示用 GMM 组合两个高斯模型来模拟图 2 中分布的示意图；

图 5 是根据本发明的方法在训练过程中的流程图；

图 6 是根据本发明的用于对样本进行分类的决策树；

图 7 是根据本发明的系统在训练部分的框图；

图 8 是根据本发明的方法在合成过程中的流程图；

图 9 表示根据本发明的动态规划的示意图；

图 10 是根据本发明的系统在合成部分的框图；

图 11 和图 12 是类属决定单元、距离计算单元及优化单元 1003 的示意框图；

图 13 表示音调树中一个树叶中的所有数据；以及

图 14 表示相邻单元之间出现不合理跳跃时的情形。

#### 具体实施方式

以下将参照附图对本发明的优选实施方式进行说明。应当指出，尽管将以简体中文为例对本发明的距离定义方式进行说明，但是本领域技术人员能够理解，本发明的实施与其所使用的语言无关，本发明同等地适用于任何语言。

GMM 刻画了当前类中样本的分布情况。分布密集的位置，其输

出概率偏大，反之输出概率偏小。单元与 GMM 模型的距离描述的是该单元与模型所在的类的近似度。GMM 作为该类的抽象表示，单元与 GMM 模型的距离可以用单元在该模型中的概率输出来表示，概率大的距离小，概率小的距离大。

G 代表 GMM 模型，单元 X 在 G 中的概率输出为  $P(X|G)$ 。单元 X 与 G 的距离定义为  $D(X, G)$ 。若存在两个单元 X1 和 X2，如果  $P(X1|G) > P(X2|G)$ ，则  $D(X1, G) < D(X2, G)$ ；如果  $P(X1|G) < P(X2|G)$ ，则  $D(X1, G) > D(X2, G)$ ；如果  $P(X1|G) = P(X2|G)$ ，则  $D(X1, G) = D(X2, G)$ 。

首先参见图 5，图 5 示出了根据本发明的方法在训练阶段的流程图。根据本发明的方法在步骤 S510 开始，然后进入步骤 S520。

步骤 S520 用于对要进行 TTS 的文本进行分析，以获得带描述性韵律注释的文本。然后，根据本发明的方法进入步骤 S530，在这里对文本中的样本进行分类，如本领域技术人员所知，“样本”可以表示建模所依据的条件，例如，若针对持续时间建立模型，则该持续时间本身即为样本。在对样本进行了分类之后，该方法进入步骤 S540，其中会对每个分类产生 GMM 模型。在产生了 GMM 模型后，根据本发明的方法的训练过程在步骤 S550 中结束。

如此产生的 GMM 模型将被用于后续的语音合成过程，如下文所述。

以下将对具体的分类方式进行详细说明。如本领域技术人员所知，可以采用多种方式对样本进行分类。例如，可以依据维度对样本进行分类；也可以依据例如“持续时间”这样的条件对样本分类。

然而，根据本发明的一种优选实施方式，通过采用决策树来对样本进行分类。与其他分类方式相比，决策树是一种数据驱动的自动分类方法，它通过数据来决定分类，从而不需要使用者具备有关分类的很多知识。在 TTS 中，决策树被普遍用于与上下文相关的分类或预测。决策树可以有多种类型，图 6 示例性地给出了决策树的典型表示。

在树的父节点中的所有数据由来自预定义的问题集的优化问题

分成两个子节点。遵循预定的准则，任何子节点中的距离都小，而两个子节点之间的距离大。在每个分离过程后，可以实现可选的功能，以在所有树叶间合并类似的节点。分离、停止分离和合并均由预定的准则优化。

如图 6 所示，假定条件 1 表示是否为句首，条件 2 表示是否为 4 声，条件 3 表示是否后跟轻声。在一个样本经过了决策树中足够多的节点（这里为 3 个）而足以得到一个适当的分类时，则为该分类生成 GMM 模型。由于对分类生成 GMM 模型的各种方式对于本领域技术人员而言是公知的，因此这里不再赘述。

此外，如果决策树中的两个分类之间的距离足够接近，可以将这两个分类合并，以用于决策树中的后续分类。如图 6 所示，该决策树的条件 1 和条件 2 的“否”分支在距离上接近（即，彼此相似），因此它们被合并后在条件 4 处用于进一步的分类。如本领域技术人员所知，根据本发明的距离定义系统可以包括一个合并单元，用于执行上述决策树中的分支合并操作。

有关生成 GMM 模型的更多信息，可参见例如 N. Kambhatla, “Local Models and Gaussian Mixture Models for Statistical Data Processing” PhD thesis, Oregon Graduate Institute of Science and Technology, 1996 年 1 月。

以下将参照图 7 对本发明的训练系统进行描述。如图 7 所示，根据本发明的训练系统 700 包括的文本分析单元 701，韵律预测单元 702 以及一个与韵律预测单元 702 相连的 GMM 模型存储单元 703，用于存储为每个分类产生的 GMM 模型。

根据本发明的优选实施方式，所述训练系统 700 还可包括用于存储一系列优化问题的单元（未示出）、针对所述各优化问题进行判断的单元（未示出）以及对适当的分类进行合并的单元（未示出），以实现前述决策树。

以下将讨论根据本发明的合成方法与系统。

图 8 示出了根据本发明的合成方法的流程图。

根据本发明的合成方法在步骤 S810 开始并随后进入步骤 S820。在步骤 S820 中，决定需要合成的单元（例如，可以是文本中包含的字）的类属，以确定该类的 GMM 模型。例如，可通过决策树中的一系列问题来决定所述类属，以在 GMM 模型库中找到与之对应的 GMM 模型。接着，在步骤 S830 中，针对该类属中的各个待选样本，计算这些样本与找到的 GMM 模型间的距离。有关计算该距离的具体方式将在下文中详细描述。在计算了每个样本的距离之后，在步骤 S840 中取出具有最小距离的那个样本作为最优的样本，以用于合成。随后，根据本发明的方法在步骤 S850 中结束。

以下将对步骤 S830 进行详细描述。如前所述，根据本发明的方法需要计算每个需要合成的单元与其 GMM 模型间的距离，且距离最小的最优。所述距离又称为目标代价（target cost）。当针对每个需要合成的单元都进行了上述计算后，将所得到的所有距离最小的需要合成单元相加，即可得到最终的合成语音。

在本发明中，采用动态规划（dynamic programming）来计算代价（cost）。即，通过本地的优化代价函数估计来找出全局优化路径。

根据本发明的一种优选实施方式，除计算上述目标代价之外，还可计算转换代价。目标代价指一个需要合成单元与其 GMM 模型之间的距离。连续的两个合成单元的语音参数需要满足一定的转换关系，只有匹配的单元才能得到较高的合成自然度，转换模型从模型的角度刻画这种转换关系。

连续的两个合成单元的语音参数的转换特性在当前转换模型中的评价，即该转换特性与当前转换模型之间的距离称为转换代价。该距离也可用 GMM 模型距离来表示。

如图 9 中的粗实线所显示的路径所示。通过累计每个节点的目标代价和转换代价，可以获得每个可能路径的代价。在评估了所有可能的路径后，以最小的代价生成全局优化的路径。

如图 9 所示，假定  $C(1, x)$  代表前述文本中的“这”， $C(2, x)$  代表前述文本中的“是”， $C(1, x)$  代表前述文本中的“一”，

等等。根据本发明的一种方式，可以只选出每个需要合成单元的最小目标代价，并直接将这些目标代价最小的单元合成起来进行语音输出。然而，根据本发明的另一种优选实施方式，可以同时考虑路径的转换代价。在图 9 中认为 C(1, 2) -C(2, m2) -C(3, 1) 被认为是目标代价加转换代价最小的路径。

本发明的合成过程可通过如图 10 所示的合成系统 1000 实现。根据本发明的合成系统 1000 包括类属决定单元 1001，用于决定需要合成的单元的类属以从 GMM 模型库中确定对应的 GMM 模型。在确定了 GMM 模型后，由距离计算单元 1002 针对各待选样本计算这些样本与找到的 GMM 模型间的距离。接下来，由优化单元 1003 对计算得到的距离进行评估，以得到距离最小的合成单元。然后，所述距离最小的需要合成的单元被输出到合成单元 1004，以形成物理声音。

此外，所述距离计算单元 1002 还可进一步包括目标代价计算单元以及转换代价计算单元（未示出）。

以上描述了基于 GMM 的距离定义。有两种典型的情况来使用所述距离定义。一种是评估给定样本与给定分类之间的距离，这是基于单元选择（unit-selection）的方法的任务，另一种是通过在给定概率分布的空间中进行搜索来预测明确的发音参数。

将所述距离定义应用于 TTS 系统中的单元选择的步骤如下：

（在训练过程中）

1. 从带标记的语料库中提取发音参数及其上下文信息；
2. 基于发音参数的上下文信息对其进行分类，并由基于 GMM 的距离定义给出发音参数之间的距离；
3. 生成 GMM 模型来描述在步骤 2 中生成的每个分类的概率分布。

（在合成过程中）

4. 从文本分析单元的结果获取每个发音段（即，要合成单元）的上下文信息；
5. 查找每个段的上下文等同类属，所述类属对应于一个 GMM

模型；

6. 通过基于 GMM 的距离定义评估段的所有候选；
7. 基于步骤 6 中给出的距离以及总体优化准则（例如，动态规划）找出总体优化候选序列；
8. 进行语音合成以生成物理语音。

将所述距离定义用于明确预测的步骤如下：

1. 从带标记的语料库中提取发音参数及其上下文信息；
2. 基于发音参数的上下文信息对其进行分类，并由基于 GMM 的距离定义给出发音参数之间的距离；
3. 生成 GMM 模型来描述在步骤 2 中生成的每个分类的概率分布。

（在合成过程中）

4. 从文本分析单元的结果获取每个发音段（即，要合成单元）的上下文信息；
5. 查找每个段的上下文等同类属，所述类属对应于一个 GMM 模型；
6. 在 GMM 混合模型序列的空间中，基于距离定义以及总体优化准则搜索最佳值，最佳值序列被认为是明确预测；
7. 根据步骤 6 中的明确预测进行合成。

为实现上述操作，所述类属决定单元 1001 可进一步包括：用于获取每个需要合成的单元的描述性韵律注释信息的韵律注释信息获取装置；用于查找每个需要合成单元的类属的查找装置，所述类属对应于一个 GMM 模型；以及用于在 GMM 混合模型序列的空间中，基于距离定义以及总体优化准则搜索最佳值，从而最佳值序列被作为 GMM 模型的明确预测的最佳值搜索装置。

相应地，所述距离计算单元 1002 可进一步包括：用于获取每个需要合成单元的描述性韵律注释信息的韵律注释信息获取装置；用于查找每个需要合成单元的属，所述类属对应于一个 GMM 模型的查找装置；以及用于通过基于 GMM 的距离定义评估需要合成单元的所有候

选的候选评估装置；以及

而所述优化单元 1003 可进一步包括用于基于评估步骤中给出的距离以及总体优化准则找出总体优化候选序列，以进行语音合成的总体优化候选序列获得装置。

图 11、12 示意地给出了类属决定单元 1001、距离计算单元 1002 及优化单元 1003 的例证构成。

应当指出，上述各可以有多种实现方式，例如采用计算机程序代码单元，或电子逻辑电路，这是本领域技术人员能够理解的，因此不再详述。

基于 GMM 的距离定义的本质是精确地模拟在 TTS 数据中定义的类属的概率分布，然后给出孤立的样本和类属之间的距离，这对于基于单元选择的方法是非常关键的。

基于 GMM 的距离定义的另一个优点在于可以在 TTS 的统计技术中顺利的部署诸如容错、适应之类的某些成熟算法。

在根据本发明的 TTS 训练与合成过程中，一种优选的实施方式是将决策树、GMM 和动态规划三者组合起来，成为一种基于单元选择的 TSS。其中，GMM 被用于描述合成序列中每个节点的目标的预测，以及相邻节点间的转换预测。

这种组合的要点在于：

- 首先，使用基于决策树的分类算法将语料库中的所有韵律矢量或段分成合理的类。类的数量取决于预定义的准则以及语料库中的数据量。

- 对于每个类，基于该类中的数据对 GMM 进行训练。

- 动态规划中的代价函数被改变为 log 概率分布，这意味着全局优化路径是具有最大累计 log 概率值的路径。它可以被看作是代价函数的负操作。

- 每个节点的韵律目标的 GMM 生成目标 log 概率函数。目标预测是某些 TTS 系统中的普遍方法。并且两个相邻节点的韵律转换的 GMM 会生成转换 log 概率分布。

以下将对韵律转换进行详细说明。

如前所述，目标韵律被广泛使用。它是预测每个段的期望和基于预测进行选择的自然方式。但最大的问题是数据分散问题。例如，图13是音调树中一个树叶中所有数据的图形表示。范围看起来很大，而且分布看起来平均。尽管容易通过GMM模型给出目标概率预测，难以期望只有目标模型就能得到好的选择结果。

有人试图通过平滑准则来解决这一问题。但是，这只能解决一些问题，而不是全部，而且最重要的问题是，有时，简单的平滑准则会使情形更差。

图14详细表示了这一现象。相邻单元之间的两个参数可能存在合理的跳跃或不合理的跳跃。跳跃的幅度值是与上下文相关的。

因此，提出转换韵律的概率模型，以对两个相邻段之间的变化进行建模。有多种与转换相关的韵律参数，例如，两个段之间的 $\log$ 音调差、 $\log$ 持续时间和音量值。很自然，转换模型在动态规划搜索方案中生成转换概率输出。

根据本发明的一种优选方式，将全新的转换韵律概率模型集成到决策树、GMM和动态规划的组合中。一方面，语库中的所有段都可用于训练目标概率预测树和单个转换概率树，这意味着在概率模型建立中不存在数据稀疏问题。由于有了转换模型，即使仍然存在数据分散问题，其影响也被部分消除，因而使得预测的韵律更稳定、更合理。

以上对本发明的优选实施方式进行了说明，但本领域技术人员应当理解，本发明给出的各优选实施方式只是为了说明的目的，不应理解为对本发明的任何限制。本发明可以采用软件、硬件或二者结合的形式实现。本领域技术人员可以根据上述描述获得有关本发明的任何变形和改进，但这些变形和改进都包括在随附权利要求书中所限定的本发明的范围和精神内。

图1

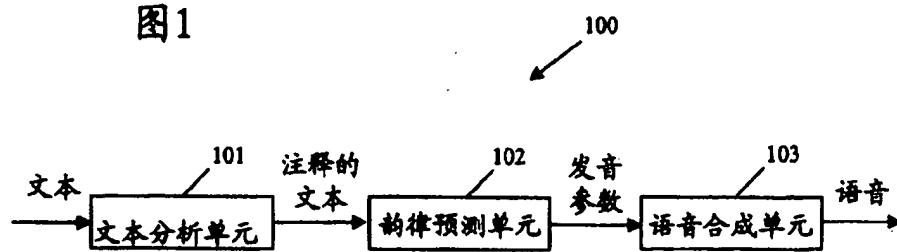


图2

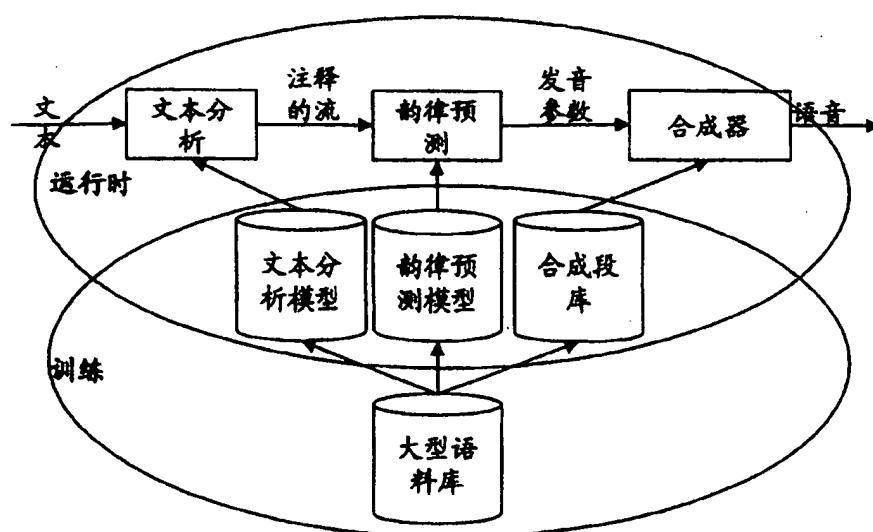


图 3

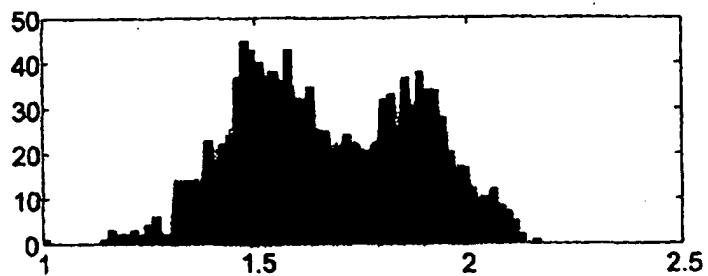


图 4



图 6

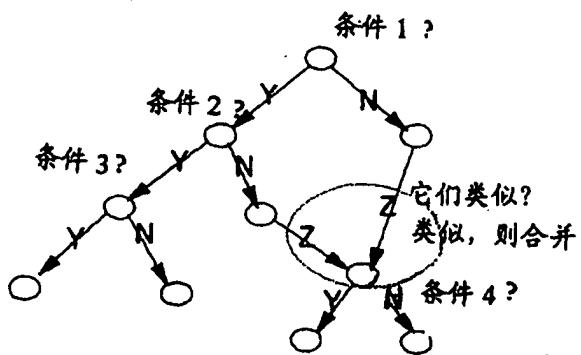


图5

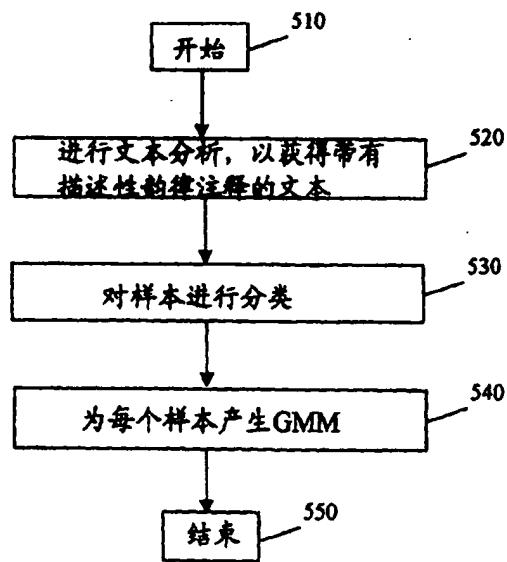


图7

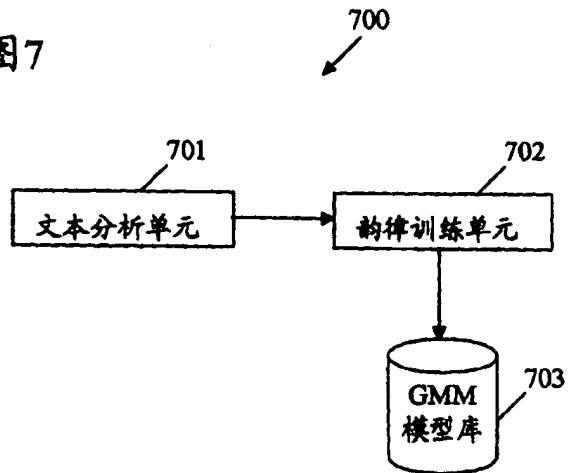


图8

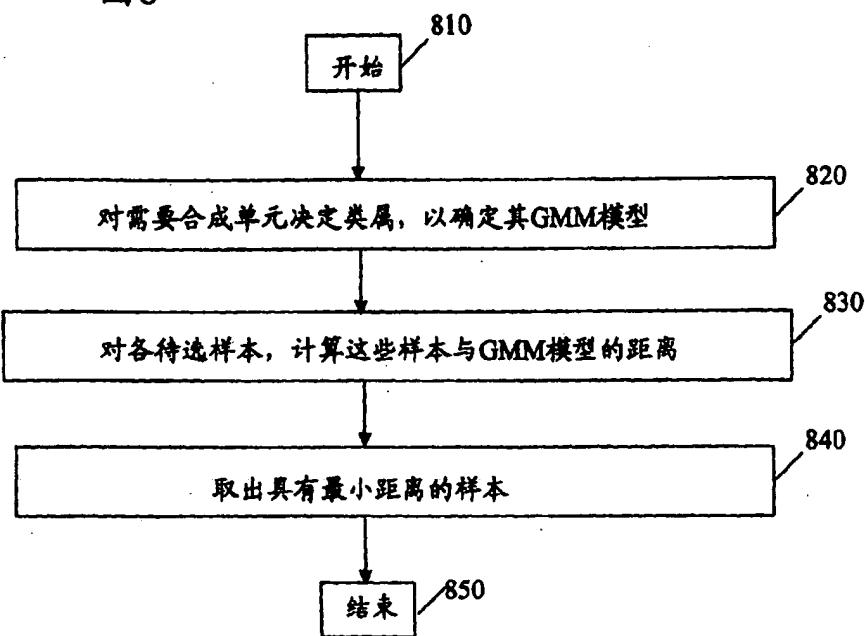


图10

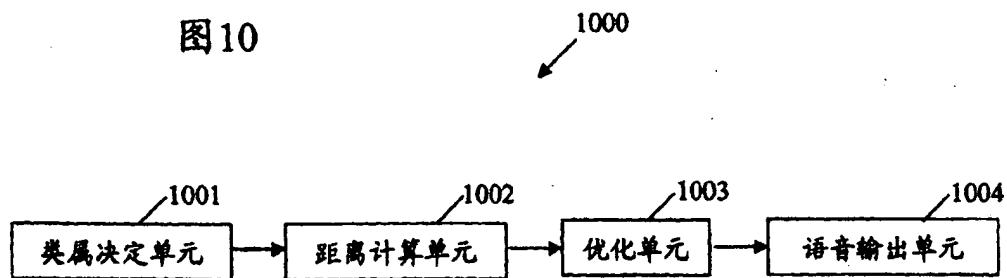


图 9

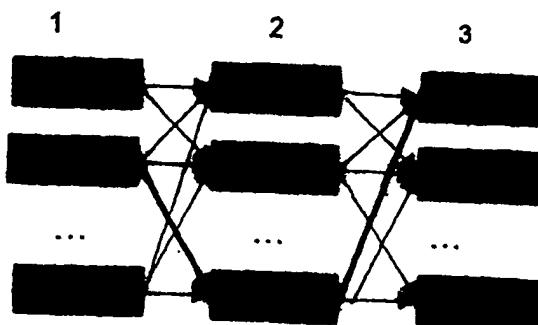


图 13

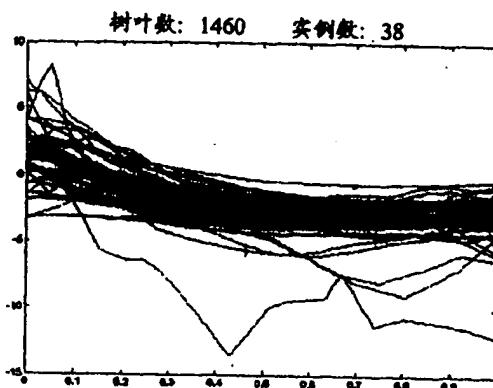
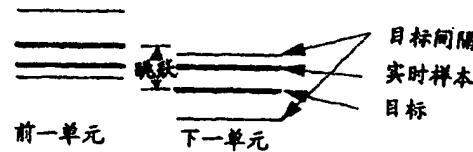
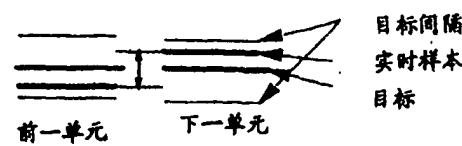


图 14



## 7.1 没有转换约束的不合理跳跃



## 7.2 简单的平滑会消除合理跳跃

图11

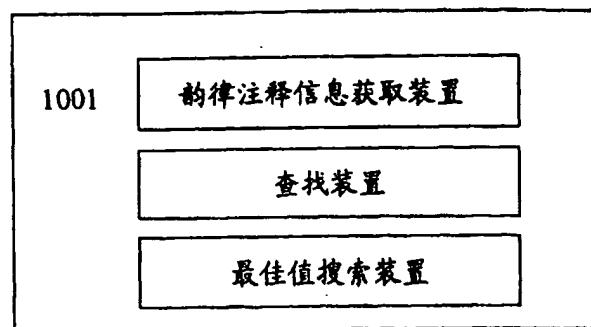


图12

